PCT/NZ99/00227

4

Intellectual Property Office of New Zealand
Te Pou Rāhui Hanga Hou

CERTIFICATE

REC'D 09 FEB 2000
WIPO PCT

This certificate is issued in support of an application for Patent registration in a country outside New Zealand pursuant to the Patents Act 1953 and the Regulations thereunder.

I hereby certify that annexed is a true copy of the Provisional Specification as filed on 23 July 1999 with an application for Letters Patent number 336906 made by THE HORTICULTURE & FOOD RESEARCH INSTITUTE OF NEW ZEALAND LTD.

Dated 31 January 2000.

Neville Harris

Neville Harris
Commissioner of Patents

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

336906

Patents Form No. 4

PATENTS ACT 1953

PROVISIONAL SPECIFICATION

SERINE PROTEASE INHIBITOR

We, **THE HORTICULTURE AND FOOD RESEARCH INSTITUTE OF NEW ZEALAND LIMITED**, a New Zealand company of Batchelar Research Centre, Highway 57, Palmerston North, New Zealand do hereby declare this invention to be described in the following statement:

(followed by page 1a)

INTELLECTUAL PROPERTY OFFICE OF N.Z.
23 JUL 1999
RECEIVED

SERINE PROTEASE INHIBITOR

This invention relates to a serine protease inhibitor. More particularly, it relates to a protein which *inter alia* exhibits anti-thrombin activity.

BACKGROUND

Thrombin is a serine protease involved in blood coagulation. It has specificity for the cleavage of arginine-lysine bonds as well as cleaving an arginine-threonine bond in pro-thrombin, releasing pre-thrombin which is subsequently cleaved to produce active thrombin. This active thrombin can then release more thrombin from pro-thrombin. In blood clotting and coagulation, thrombin cleaves fibrinopeptide B from fibrinogen as well as converting blood factors IX to IXa, V to Va, VIII to VIIIa and XIII to XIIIa.

Inhibitors of thrombin therefore inhibit coagulation and have application in any procedure where coagulation is undesirable. One such application is in the collection and storage of blood products. Another is in medicaments for preventing or reducing coagulation for example in treating or preventing cardiac malfunctions.

Anti-thrombin agents are known. One example is anti-thrombin III (AT-III). However, AT-III is capable of effectively inhibiting thrombin only in the presence of heparin.

The applicants have now identified a novel protein which has anti-thrombin activity and which does not require heparin as a cofactor. It is towards this protein that the present invention is broadly directed.

SUMMARY OF THE INVENTION

Accordingly, in a first aspect the present invention provides a protein obtainable from *Perna canaliculus* which has an approximate molecular weight of 75 kDa determined by PAGE and which has anti-thrombin activity, or an active fragment thereof. The molecular weight of the protein inferred from its corresponding nucleotide encoding sequence is about 55 kDa. The protein is referred to as "pernin".

Conveniently, the protein is obtainable from the haemolymph of *P. canaliculus*.

Preferably, the protein is a self-aggregating protein.

More preferably, the protein includes one or more of the following amino acid sequences:

(a) DGEQCNDGQN
(b) QGGHEVESERVACCVIGRA
(c) GQSHPEIVH
(d) YHGHDDA
(e) VVNEVHH.

Conveniently, sequence (a) is at or towards the N-terminal end of the protein.

Most conveniently, the protein comprises the following amino acid sequence:

```
D G E Q C N D G Q N K D D H H D D H H D D H H D D H D D D D
E T M H Y A Q C E M E P N P H M A S S L H H H V H G S I E L
S Q K G H G A V Y L E L H L V G F N T S E D H D D H H H G L
H L H M L G D M S A G C D S I G E L Y N A H P E K H A D P G
D L G D L V D D D R G V V N E V H H Y A W L D I D G T A P N
T E A L I G H S M T I L Q G S H T D A D T P A S R I A C C V
I G H G K A R P E T A A A L H H E L E E D K T E H Y A H C D
V R S N T H Q P K A L H H H V H G T I D F K Q V G Y G D L E
V S Y H L E G F N V S D D H K D H L H D V Q I Y A N G D L T
S G C D N L G A K Y D P H E D Y H S E L G D L G D I H D D D
H G V V N E S H R Y S W I N I F G D D S V L G R S I A I H Q
R D H L H K S A K I A C C V I G R G Q S H P E I V H R A K C
V V R P N T E S T G L H H H V S G S I T F E Q T P G G S T H
M T A D L K G F N V S E D L S H H R H G V Q L H E W G D M S
H G C H S L G R M Y H G H D D A H D P K R P G D L G D V I D
D S H G I V H S T R T F D H L N V E D L N A R S L V I M Q G
G H E V E S E R V A C C V I G R A
```

In a further aspect, the invention provides a polynucleotide molecule encoding the protein defined above or an active fragment thereof. Usually, the polynucleotide molecule will be DNA.

In one embodiment, said polynucleotide comprises the nucleotide sequence:

5'GATGGGGAGCAGTGTAACGATGGGCAGAACAAAGATGACCACCATGACGA
CCACCACGATGATCACCATGACGACCATGATGATGATGATGAAACAATGCACT
ATGCCCAGTGTGAAATGGAACCAAACCCTCATATGGCTAGCAGCCTTCACCA
CCATGTCCATGGCAGCATAGAGTTGTCACAGAAGGGTCATGGAGCTGTTTAT
CTAGAACTTCATCTTGTCGGATTCAACACAAGTGAAGACCATGACGACCACCA
TCATGGACTTCATCTGCACATGCTTGGTGACATGTCAGCAGGTTGTGATTCTA
TTGGCGAACTGTACAATGCTCACCCAGAAAAACATGCTGACCCTGGTGACCT
CGGTGACCTGGTTGACGATGATAGGGGCGTGGTTAATGAAGTTCATCATTATG
CTTGGTTGGACATTGATGGTACAGCACCAAACACCGAAGCTCTCATTGGACA
CTCAATGACTATTTTACAAGGGAGTCACACCGATGCTGATACCCCAGCCAGTA
GAATCGCCTGTTGTGTTATTGGTCATGGAAAAGCTCGCCCAGAAACAGCAGC
TGCTCTACATCACGAGCTAGAGGAAGATAAAACTGAGCATTATGCCCATTGTG
ACGTAAGATCTAATACACACCAACCAAAGGCTCTTCATCATCATGTCCACGGA
ACCATCGATTTCAAACAAGTTGGTTATGGTGACCTTGAAGTGTCCTACCATTTA
GAGGGATTTAATGTAAGTGATGACCACAAAGATCATCTCCATGACGTACAGAT
CTACGCCAACGGTGACCTGACCAGTGGATGTGATAACCTCGGTGCTAAATAT
GATCCTCATGAAGATTACCACAGTGAGTTGGGTGATCTAGGAGATATTCACGA
TGATGACCATGGCGTTGTCAATGAAAGCCACAGATATTCCTGGATCAATATCT
TCGGTGATGACAGTGTCCTGGGACGTTCTATTGCCATTCACCAAAGAGACCAT
CTTCATAAAAGTGCCAAAATTGCCTGTTGTGTCATAGGACGTGGACAGAGCCA
TCCAGAAATTGTTCACAGAGCTAAATGTGTTGTCAGACCTAATACAGAATCTAC
TGGTTTACATCACCATGTCTCTGGTTCTATAACATTCGAACAGACCCCTGGAG
GATCAACACATATGACGGCTGATCTCAAAGGATTTAACGTTAGTGAGGACTTG
TCACATCATCGTCATGGTGTGCAGCTCCATGAATGGGGAGATATGTCCCATG
GCTGTCACTCCTTAGGCAGAATGTACCATGGTCATGATGATGCTCATGACCCC
AAAAGACCTGGTGACCTTGGTGATGTTATAGATGATTCCCATGGCATCGTTCA
TTCAACTAGAACCTTTGATCATCTTAATGTTGAAGATCTTAACGCACGTTCCCT
TGTGATTATGCAGGGCGGACATGAGGTCGAGAGTGAGAGGGTTGCTTGCTGT
GTTATAGGACGGGCA

or a variant thereof.

In still a further aspect, the invention provides a composition which includes a protein as defined above or an active fragment thereof.

Conveniently, the composition is a medicament.

Alternatively, the composition is a dietary supplement, a bioremediation agent or an oxygen carrier.

It is particularly preferred that the protein of the invention is obtained from *P. canaliculus*, more preferably obtained from the haemolymph and then purified by centrifugation.

DESCRIPTION OF THE DRAWINGS

While the present invention is broadly as defined above, it also includes embodiments of which the following description provides examples. In particular, a better understanding of the present invention will be gained through reference to the accompanying drawings in which

Figure 1: Purification of pernin from mussel haemolymph

a) light-scattering band following centrifugation of *P. canaliculus* haemolymph in CsCl; haemolymph was first centrifuged at low speed to remove haemocytes and then at high speed; the re-suspended pellet was then centrifuged in CsCl.

b) UV absorption profile (254 nm wavelength) from fractionation of the CsCl gradient; the light-scattering material in figure 1a appears as a peak.

c) protein composition in 1 ml fractions of a CsCl gradient following electrophoresis in a 12% polyacrylamide gel; the heavily stained (Coomassie) bands coincide with the position of the light-scattering and UV-absorbing regions of the gradient; the molecular weight was approximately 75 kDa as compared with polypeptide molecular weight standards (lane 6) (refer Figure 4a for standards). Lanes 1-5 and 7-9 contained samples from the CsCl gradient.

Figure 2: Virus-like particles observed by transmission electron microscopy of material in light scattering band in a CsCl gradient. Bar in micrograph represents 100 nm.

Figure 3: HPLC elution profile of pernin at 280 nm wavelength purified by CsCl gradient centrifugation..

Figure 4: SDS-PAGE profiles (12% gels) of aggregating protein species from *P. canaliculus* and other shellfish species

a) proteins extracted from whole shellfish and purified as described in Materials and Methods: lane 1: molecular weight standards (Bio-Rad, USA) :**pb** phosphorylase B, 97.4 kDa; **bsa** bovine serum albumin, 66 kDa; **ova** ovalbumin, 45 kDa; **ca** carbonic anhydrase, 31 kDa; lane 2: Greenshell™ mussel *P. canaliculus*; lane 3: blue mussel *Mytilis edulis*; lane 4: oyster *Crassostrea gigas*; lane 5: pipis *Paphies australis.*

b) PAGE analysis of human transferrin (Sigma, USA, MW *ca.* 80 kDa), a glycosylated protein, and pernin from *P. canaliculus* following treatment with endoglycosidase-F: lane 1: untreated transferrin; lane 2: transferrin treated with glycosidase-F; lane 3: untreated pernin lane 4: pernin treated with glycosidase-F.

Figure 5: Activity of *P. canaliculus* haemolymph protein following centrifugation in a 30 kDa molecular weight exclusion filter for 10 min at 1000 *g* (Ultrafree-MC filter, 30,000 MW exclusion, Millipore, USA)

a) SDS-PAGE profile of haemolymph protein at various stages of purification. Lane 1: "crude" haemolymph (haemocytes removed); lane 2: resuspended pellet after ultracentrifugation of "crude" haemolymph for 80 min at 250,000 *g*; lane 3: pernin retentate; lane 4: filtrate (no proteins evident); lane 5: molecular weight markers, (refer Figure 4a); lanes 6,7: 10-fold dilutions of samples from lanes 2 and 3.

b) Anti-thrombin activity of 30,000 MW exclusion filter retentate and filtrate.

con+ = the standard 1/41 dilution of human plasma (i.e. standard anti-thrombin III activity);

con - thrombin with no added plasma (buffer control); **filtrate**: material passed through a 30,000 MW exclusion filter;

retentate: pernin protein retained by exclusion filter.

DESCRIPTION OF THE INVENTION

As broadly outlined above, the present invention provides a novel protein having, *inter alia*, anti-thrombin activity. The protein of the invention is a protein, having an apparent molecular weight of 75 kDa, calculated by polyacrylamide gel electrophoresis (PAGE). The molecular weight inferred from the gene sequence is approximately 55 kDa.

The protein of the invention was initially identified as an extract from the New Zealand green lipped mussel *P. canaliculus*. It is therefore obtainable by extraction directly from *P. canaliculus*.

The protein of the invention can include its entire native amino acid sequence or can include only parts of that sequence where such parts constitute fragments which remain biologically active (active fragments). Such activity will normally be as a serine protease inhibitor, but is not restricted to this.

The invention also encompasses variants of the above protein. As used herein, the term "variant" covers any sequence which exhibits at least about 50%, more preferably at least 70% and more preferably yet at least about 90% identity to the sequence of the protein of the present invention. Most preferably, a "variant" is any sequence which has at least a 99% probability of being the same as the sequence of the invention. The probability of identity for protein sequences is measured by the computer algorithm BLASTP (Altschul, S F *et al*, *Nucleic Acids Res.* 25:3389-3402 (1997)). The term "variants" thus encompasses sequences where the probability of finding a match by chance is less than about 1% as measured by the above tests.

The protein of the invention together with its active fragments and other variants may be generated by synthetic or recombinant means. Synthetic polypeptides having fewer than about 100 amino acids, and generally fewer than about 50 amino

acids, may be generated by techniques well known to those of ordinary skill in the art. For example, such peptides may be synthesised using any of the commercially available solid-phase techniques such as the Merryfield solid phase synthesis method, where amino acids are sequentially added to a growing amino acid chain (see Merryfield, J. Am. Chem. Soc 85: 2146-2149 (1963)). Equipment for automative synthesis of peptides is commercially available from suppliers such as Perkin Elmer/Applied Biosystems, Inc. and may be operated according to the manufacturers instructions.

The protein, or a fragment or variant thereof, may also be produced recombinantly by inserting a polynucleotide (usually DNA) sequence that encodes the protein into an expression vector and expressing the protein in an appropriate host. Any of a variety of expression vectors known to those of ordinary skill in the art may be employed. Expression may be achieved in any appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule which encodes the recombinant protein. Suitable host cells includes procaryotes, yeasts and higher eukaryotic cells. Preferably, the host cells employed are *E. coli*, yeasts or a mammalian cell line such as COS or CHO, or an insect cell line, such as SF9, using a baculovirus expression vector. The DNA sequence expressed in this matter may encode the naturally occurring protein, fragments of the naturally occurring protein or variants thereof.

DNA sequences encoding the protein or fragments may be obtained by screening an appropriate *P. canaliculus* cDNA or genomic DNA library for DNA sequences that hybridise to degenerate oligonucleotides derived from partial amino acid sequences of the protein. Suitable degenerate oligonucleotides may be designed and synthesised by standard techniques and the screen may be performed as described, for example, in Maniatis *et al.* Molecular Cloning - A Laboratory Manual, Cold Spring Harbour Laboratories, Cold Spring Harbour, NY (1989). The polymerase chain reaction (PCR) may be employed to isolate a nucleic acid probe from genomic DNA, a cDNA or genomic DNA library. The library screen may then be performed using the isolated probe.

Variants of the protein may be prepared using standard mutagenesis techniques such as oligonucleotide-directed site specific mutagenesis.

A specific polynucleotide of the invention has the following nucleotide sequence:

5'GATGGGGAGCAGTGTAACGATGGGCAGAACAAAGATGACCACCATGACGA
CCACCACGATGATCACCATGACGACCATGATGATGATGATGAAACAATGCACT
ATGCCCAGTGTGAAATGGAACCAAACCCTCATATGGCTAGCAGCCTTCACCA
CCATGTCCATGGCAGCATAGAGTTGTCACAGAAGGGTCATGGAGCTGTTTAT
CTAGAACTTCATCTTGTCGGATTCAACACAAGTGAAGACCATGACGACCACCA
TCATGGACTTCATCTGCACATGCTTGGTGACATGTCAGCAGGTTGTGATTCTA
TTGGCGAACTGTACAATGCTCACCCAGAAAAACATGCTGACCCTGGTGACCT
CGGTGACCTGGTTGACGATGATAGGGGCGTGGTTAATGAAGTTCATCATTATG
CTTGGTTGGACATTGATGGTACAGCACCAAACACCGAAGCTCTCATTGGACA
CTCAATGACTATTTTACAAGGGAGTCACACCGATGCTGATACCCCAGCCAGTA
GAATCGCCTGTTGTGTTATTGGTCATGGAAAAGCTCGCCCAGAAACAGCAGC
TGCTCTACATCACGAGCTAGAGGAAGATAAAACTGAGCATTATGCCCATTGTG
ACGTAAGATCTAATACACACCAACCAAAGGCTCTTCATCATCATGTCCACGGA
ACCATCGATTTCAAACAAGTTGGTTATGGTGACCTTGAAGTGTCCTACCATTTA
GAGGGATTTAATGTAAGTGATGACCACAAAGATCATCTCCATGACGTACAGAT
CTACGCCAACGGTGACCTGACCAGTGGATGTGATAACCTCGGTGCTAAATAT
GATCCTCATGAAGATTACCACAGTGAGTTGGGTGATCTAGGAGATATTCACGA
TGATGACCATGGCGTTGTCAATGAAAGCCACAGATATTCCTGGATCAATATCT
TCGGTGATGACAGTGTCCTGGGACGTTCTATTGCCATTCACCAAAGAGACCAT
CTTCATAAAAGTGCCAAAATTGCCTGTTGTGTCATAGGACGTGGACAGAGCCA
TCCAGAAATTGTTCACAGAGCTAAATGTGTTGTCAGACCTAATACAGAATCTAC
TGGTTTACATCACCATGTCTCTGGTTCTATAACATTCGAACAGACCCCTGGAG
GATCAACACATATGACGGCTGATCTCAAAGGATTTAACGTTAGTGAGGACTTG
TCACATCATCGTCATGGTGTGCAGCTCCATGAATGGGGAGATATGTCCCATG
GCTGTCACTCCTTAGGCAGAATGTACCATGGTCATGATGATGCTCATGACCCC
AAAAGACCTGGTGACCTTGGTGATGTTATAGATGATTCCCATGGCATCGTTCA
TTCAACTAGAACCTTTGATCATCTTAATGTTGAAGATCTTAACGCACGTTCCCT
TGTGATTATGCAGGGCGGACATGAGGTCGAGAGTGAGAGGGTTGCTTGCTGT
GTTATAGGACGGGCA

A further polynucleotide has the sequence:

5'GATGGGGAGCAGTGTAACGATGGGCAGAACAAAGATGACCACCATGACGA
CCACCACGATGATCACCATGACGACCATGATGATGATGATGAAACAATGCACT
ATGCCCAGTGTGAAATGGAACCAAACCCTCATATGGCTAGCAGCCTTCACCA
CCATGTCCATGGCAGCATAGAGTTGTCACAGAAGGGTCATGGAGCTGTTTAT
CTAGAACTTCATCTTGTCGGATTCAACACAAGTGAAGACCATGACGACCACCA

TCATGGACTTCATCTGCACATGCTTGGTGACATGTCAGCAGGTTGTGATTCTA
TTGGCGAACTGTACAATGCTCACCCAGAAAAACATGCTGACCCTGGTGACCT
CGGTGACCTGGTTGACGATGATAGGGGCGTGGTTAATGAAGTTCATCATTATG
CTTGGTTGGACATTGATGGTACAGCACCAAACACCGAAGCTCTCATTGGACA
CTCAATGACTATTTTACAAGGGAGTCACACCGATGCTGATACCCCAGCCAGTA
GAATCGCCTGTTGTGTTATTGGTCATGGAAAAGCTCGCCCAGAAACAGCAGC
TGCTCTACATCACGAGCTAGAGGAAGATAAAACTGAGCATTATGCCCATTGTG
ACGTAAGATCTAATACACACCAACCAAAGGCTCTTCATCATCATGTCCACGGA
ACCATCGATTTCAAACAAGTTGGTTATGGTGACCTTGAAGTGTCCTACCATTTA
GAGGGATTTAATGTAAGTGATGACCACAAAGATCATCTCCATGACGTACAGAT
CTACGCCAACGGTGACCTGACCAGTGGATGTGATAACCTCGGTGCTAAATAT
GATCCTCATGAAGATTACCACAGTGAGTTGGGTGATCTAGGAGATATTCACGA
TGATGACCATGGCGTTGTCAATGAAAGCCACAGATATTCCTGGATCAATATCT
TCGGTGATGACAGTGTCCTGGGACGTTCTATTGCCATTCACCAAAGAGACCAT
CTTCATAAAAGTGCCAAAATTGCCTGTTGTGTCATAGGACGTGGACAGAGCCA
TCCAGAAATTGTTCACAGAGCTAAATGTGTTGTCAGACCTAATACAGAATCTAC
TGGTTTACATCACCATGTCTCTGGTTCTATAACATTCGAACAGACCCCTGGAG
GATCAACACATATGACGGCTGATCTCAAAGGATTTAACGTTAGTGAGGACTTG
TCACATCATCGTCATGGTGTGCAGCTCCATGAATGGGGAGATATGTCCCATG
GCTGTCACTCCTTAGGCAGAATGTACCATGGTCATGATGATGCTCATGACCCC
AAAAGACCTGGTGACCTTGGTGATGTTATAGATGATTCCCATGGCATCGTTCA
TTCAACTAGAACCTTTGATCATCTTAATGTTGAAGATCTTAACGCACGTTCCCT
TGTGATTATGCAGGGCGGACATGAGGTCGAGAGTGAGAGGGTTGCTTGCTGT
GTTATAGGACGGGCA**TGA**ATAACCTCACTAGAGTGACTTTGTCTAACATGACA
ATTAACAATTGTATAACTTCGCTAAAA**AATAAAA**CAATGACACAATGNAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAA3'

with TGA being the opal stop codon and AATAAA the polyadenylation signal.

Variants or homologues of the above sequences also form part of the present invention. Polynucleotide sequences may be aligned, and percentage of identical nucleotides in a specified region may be determined against another sequence, using computer algorithms that are publicly available. Two exemplary algorithms for aligning and identifying the similarity of polynucleotide sequences are the BLASTN and FASTA algorithms. The BLASTN software is available on the NCBI anonymous FTP server (ftp://ncbi.nlm.nih.gov) under /blast/executables/. The BLASTN algorithm version 2.0.4 [Feb-24-1998], set to the default parameters

described in the documentation and distributed with the algorithm, is preferred for use in the determination of variants according to the present invention. The use of the BLAST family of algorithms, including BLASTN, is described at NCBI's website at URL http://www.ncbi.nlm.nih.gov/BLAST/newblast.html and in the publication of Altschul, Stephen F, *et al* (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402. The computer algorithm FASTA is available on the Internet at the ftp site ftp://ftp.virginia.edu.pub/fasta/. Version 2.0u4, February 1996, set to the default parameters described in the documentation and distributed with the algorithm, is preferred for use in the determination of variants according to the present invention. The use of the FASTA algorithm is described in the W R Pearson and D.J. Lipman, "Improved Tools for Biological Sequence Analysis," *Proc. Natl. Acad. Sci. USA 85*:2444-2448 (1988) and W.R. Pearson, "Rapid and Sensitive Sequence Comparison with FASTP and FASTA," *Methods in Enzymology 183*:63-98 (1990).

All sequences identified as above qualify as "variants" as that term is used herein.

While the above synthetic or recombinant approaches can be taken to produce the protein of the invention, it is however practicable (and indeed presently preferred) to obtain the protein by isolation from *P. canaliculus*. This reflects the applicants' finding that the protein is the dominant protein of the haemolymph of *P. canaliculus* and also that the protein is self-aggregating. It can therefore be isolated in commercially significant quantities direct from the mussel itself. For example, approximately 2 mg of the protein can be obtained per ml of haemolymph.

Once obtained, the protein is readily purified if desired. This will generally involve centrifugation in which the self-aggregating nature of the protein is important. Other approaches to purification (eg. chromatography) can however also be followed.

Furthermore, if viewed as desirable, additional purification steps can be employed using approaches which are standard in this art. These approaches are fully able to deliver a highly pure preparation of the protein.

Once obtained, the protein can be formulated into a composition. The composition can be, for example, a therapeutic composition for application as a pharmaceutical, or can be a health or dietary supplement. Again, standard approaches can be taken in formulating such compositions.

The invention will now be described more fully in the following experimental section which is provided for illustrative purposes only.

EXPERIMENTAL

A. Materials and Methods

A.1 Shellfish: *Perna canaliculus* (the New Zealand green-lipped mussel; the Greenshell™ mussel) were obtained at retail supermarket outlets or from mussel farmers directly; other shellfish species were obtained from retail outlets except for the blue mussel *Mytilis edulis* which was supplied by Sanford's Fisheries (Havelock, New Zealand).

A.2 Extracts: Mussel extracts were prepared by homogenising whole, shucked mussels (up to 120 mm length) in a commercial food processor with the addition of 0.02 M sodium phosphate buffer, pH 7.2. Dichloromethane (1/2 volume) was mixed with the aqueous extract, centrifuged at low speed (6000 rpm, GSA rotor, Sorvall RC-5B centrifuge at 4 °C). Polyethylene glycol (PEG) (MW 6000) was added to the aqueous phase to a final concentration of 10% (w/v) and NaCl to 0.5 M and stirred at 4-6 °C overnight. Following low speed centrifugation the PEG-precipitate was resuspended in approximately 1/10 volume of sodium phosphate buffer. After another cycle of low-speed centrifugation the supernatant was centrifuged at high speed (50,000 rpm in a Beckman 60Ti rotor at 4 °C for 60-80 minutes). The resultant pellet was resuspended in a small volume of phosphate buffer and clarified by low speed centrifugation.

A.3 Polyacrylamide gel electrophoresis: 12% polyacrylamide gels (8 x10 cm; 1 mm thick) were cast using a prepared stock solution according to the manufacturer's instructions (40% acrylamide/bis solution 37.5:1, Bio-Rad, USA); commercially available 12% gels (Bio-Rad, USA) were also used. Samples (10 μl) were applied to lanes and the gels run at 160 V using a standard Tris/Glycine/SDS buffer (Bio-Rad, catalogue 161-0732) until the bromphenol blue marker reached the bottom of the gel. Gels were stained with BM Fast Stain Coomassie® (Boehringer Mannheim, Germany) and destained as per the manufacturer's instructions.

A.4 **Glycosylation test:** Samples were treated with N-glycosidase F (PNGase F from *Flavobacterium meningosepticum*; Boehringer Mannheim Biochemica, Germany) according to the manufacturer's directions. Treated and untreated samples were run in a standard 12% polyacrylamide gel.

A.5 **Thrombin inhibition assay:** Kinetic assays were done using an Accucolor™ Antithrombin III kit (catalogue no. CRS105, Sigma Diagnostics, USA) with the reagents prepared according to the supplier's directions. Standard plasma was supplied by Instrumentation Laboratories (Italy) and used at the recommended dilution of 1/41. Samples of purified mussel protein in water were diluted 9/10 by adding 10X Sigma sample buffer. Heparin was purchased from Instrumentation Laboratories. Thrombin activity was estimated colorimetrically at 405 nm using a chromogenic substrate (H-D-HHT-L-Ala-L-Arg-pNa.2AcOH, catalogue no. A 8058, Sigma, USA) and a Multiskan Biochromatic plate reader (Labsystems, Finland)

A.6 **Isopycnic gradients:** CsCl (Boehringer Mannhein, Germany) solutions were prepared in 0.1 M sodium phosphate buffer, pH 7.2 and filtered through a 0.22 μm membrane (Acrodisc, Gelman Sciences, USA) to clarify. Two step gradients (1.25 g/cc top layer containing the sample and 1.45 g/cc bottom layer) were prepared as described by Scotti (1985) and centrifuged for approximately 17 hours at 20 °C in a Beckman 70Ti rotor at 30,000 rpm. The resultant gradient was fractionated by inserting a 100 μl glass capillary tube into the gradient and slowly pumping out the contents. UV absorbance was monitored by passing through a Uvicord spectrophotometer (LKB Produkter, Sweden). Fractions were collected and the refractive indices measured using an Abbé refractometer (Bellingham and Stanley, UK) and the density estimated using regression equations according to the method of Scotti (1985).

A.7 **Porous glass chromatography:** Controlled pore glass (CPG 240-80, Sigma Chemical Co., USA) was treated according to the suppliers directions. A 1 cm x 100 cm column (Bio-Rad, USA) was prepared. Samples (1-2 ml) were loaded onto the column and eluted with 0.1 M sodium phosphate buffer, pH 7.2, through a Uvicord spectrophotometer, fractions being collected at regular intervals.

A.8 **Estimation of protein concentration:** Concentrations were estimated using a bovine serum albumin standard (Blot Qualified BSA, Promega, USA) by UV absorption according to the method of Layne (1957) using the equation: mg/ml protein = 1.55^*A_{280} - 0.76^*A_{260}. Alternatively, concentration was estimated by the Bradford reaction using reagent supplied by Bio-Rad (USA) at a wavelength of 620 nm..

A.9 **High performance liquid chromatography:** Reversed-phase HPLC was performed on an HP 1050 Ti-series HPLC (Hewlett Packard, USA) fitted with an analytical 300 Å Vydac C-18 column, 25 cm x 4.6 mm i.d.. The 10 µl sample in water was eluted with a 0-100% acetonitrile in water (v/v) gradient over 60 min and the absorption at 218 and 280 nm was recorded.

B. Results

A light-scattering band was seen after centrifugation of extracts of whole Greenshell™ mussels in CsCl gradients **(Figures 1a and 1b)**. The density of this band was estimated at 1.368 g/cc. A minor band was sometimes observed at approximately 1.390 g/cc. If rebanded in CsCl the 1.390 band yielded two bands - one at 1.390 g/cc and a second at 1.368 g/cc. SDS-PAGE analysis of fractions of either density gave similar polypeptide profiles with a single major band. The molecular weight of the protein by PAGE was estimated as 75,000 (75 kDa) **(Figure 1c)**. Several minor bands of higher molecular weight and an additional minor band of 45 kDa were also seen. The main band (called pernin) at 75 kDa was always at great excess compared to the minor bands. When material from the light-scattering material from CsCl gradients were examined by electron microscopy, particles resembling those of "empty" small RNA viruses were seen **(Figure 2)**. However a UV wavelength scan (data not shown) indicated that little, if any, nucleic acid was present and that the particles were mainly composed of protein. HPLC showed the CsCl band to be composed almost solely of a single species of protein **(Figure 3)**. Since HPLC indicated a high degree of purity, the higher molecular weight polypeptides are presumed to be multimers of pernin. It is likely that the minor, lower molecular weight band is degraded pernin.

Chromatography, on a CPG 240-80 column, of semi-purified extracts, or of material banded in CsCl, showed that the majority of pernin was eluted in the exclusion volume using low molarity phosphate or Tris buffer as the eluent. In contrast, a protein of similar size, bovine serum albumin (68 kDa), was included in the column matrix. It appears, therefore, that pernin does aggregate into large, particle-like structures under certain conditions as suspected from the particles seen in **Figure 2**. HPLC confirmed that pernin from *P. canaliculus* obtained by CPG chromatography was highly purified. Aggregating protein species were also detected in extracts of other shellfish: the blue mussel *Mytilis edulis*, the oyster *Crassostrea gigas*, and New Zealand pipis *Paphies australis* but not in scallops *Pecten novaezealandiae*. These polypeptides were lower in molecular weight than pernin **(Figure 4a)**. The pernin from *P. canaliculus* is N-glycosylated as shown by a reduction in molecular weight when treated with endoglycosidase-F before PAGE **(Figure 4b)**.

The yield of pernin from whole mussel extractions averaged about 200 μg/mussel. Improved yields of pernin were obtained by extracting haemolymph directly from live *P. canaliculus*. A small notch was made in the shell using a triangular file and a 30 gauge needle inserted into the posterior adductor muscle. From 1 to 5 ml of haemolymph can be withdrawn easily. The haemolymph was spun at low speed (≈1000 *g*) to remove haemocytes and the resulting supernatant processed by ultracentrifugation, for example at 250,000 *g* for 40 minutes, followed by either CPG chromatography eluting with 0.1 M sodium phosphate buffer, pH 7.2, or isopycnic banding in CsCl in phosphate buffer. The pernin obtained in this way appeared no different than that purified from whole mussels and had the advantage of a 30-fold average increase in yield from each mussel. Haemolymph contained around 2 mg/ml (average ≈5-6 mg/mussel) of pernin which is by far the most predominant polypeptide species (**Figure 5a**). The time to purify pernin was reduced from about 5 days to 1 day.

Microsequencing of the N-terminal region and internal fragments generated by chemical and enzymatic cleavage from purified pernin was performed and generated the following sequences of cleavage fragments:

(a) DGEQCNDGQN

(b) QGGHEVESERVACCVIGRA

(c) GQSHPEIVH

(d) YHGHDDA

(e) VVNEVHH.

These sequences code for amino acids as follows:

CODE:	
A	alanine
C	cystine
D	aspartic acid
E	glutamic acid
F	phenylalanine
G	glycine
H	histidine
I	isoleucine
K	lysine
L	leucine
M	methionine
N	asparagine
P	proline
Q	glutamine
R	arginine
S	serine
T	threonine
V	valine
W	tryptophan
Y	tyrosine

The sequence data was then compared with amino acid sequences in searchable computer data bases. Some sequences were found to be of particular interest:

a) a 10 amino acid residue sequence from the N-terminus of pernin (sequence (a) above) showed <u>only</u> homology with an 8 base anti-thrombin protein sequence from terrestrial leeches (data from US Patent 5,455,181 Oct 3, 1995: sequence 10).

Perna canaliculus pernin	**2** GEQCNDGQ **9**	
matching amino acids	*G+ CNDGQ*	
leech anti-thrombin	**5** GQSCNDGQ **12**	

identities: 6/8 (75%) positives: 7/8 (87%);
"+" indicates an equivalent amino acid;
the bolded numerals indicate amino acid position

b) An internal cleavage product (sequence (b) above) was shown to be have homology to the Cu-Zn class of proteins known as "SODs" (superoxide dismutases).

Each of fragments (a) to (e) are part of the larger pernin amino acid sequence:

1	**DGEQCNDGQN**	KDDHHDDHHD	DHHDDHDDDD	ETMHYAQCEM	EPNPHMASSL
51	HHHVHGSIEL	SQKGHGAVYL	ELHLVGFNTS	EDHDDHHHGL	HLHMLGDMSA
101	GCDSIGELYN	AHPEKHADPG	DLGDLVDDDR	G**VVNEVHH**YA	WLDIDGTAPN
151	TEALIGHSMT	ILQGSHTDAD	TPASRIACCV	IGHGKARPET	AAALHHELEE
201	DKTEHYAHCD	VRSNTHQPKA	LHHHVHGTID	FKQVGYGDLE	VSYHLEGFNV
251	SDDHKDHLHD	VQIYANGDLT	SGCDNLGAKY	DPHEDYHSEL	GDLGDIHDDD
301	HGVVNESHRY	SWINIFGDDS	VLGRSIAIHQ	RDHLHKSAKI	ACCVIGR**GQS**
351	**HPEIVH**RAKC	VVRPNTESTG	LHHHVSGSIT	FEQTPGGSTH	MTADLKGFNV
401	SEDLSHHRHG	VQLHEWGDMS	HGCHSLGRMY	**HGHDDA**HDPK	RPGDLGDVID
451	DSHGIVHSTR	TFDHLNVEDL	NARSLVIMQG	**GHEVESERVA**	**CCVIGRA**

(Bold characters indicate directly sequenced fragments (a) to (e)).

Anti-thrombin Activity

The possibility that pernin could function as an anti-thrombin agent was examined in a kinetic assay for thrombin inhibition which was performed in our laboratory as described above. This verified that pernin had inhibitory activity. When a purified preparation of pernin was centrifuged through a 30,000 MW exclusion filter **(Figure 5a)**, all the anti-thrombin activity was in the retentate and no detectable activity was present in the filtrate **(Figure 5b)**. The standard serum was diluted 1/41 as recommended for this assay system; the pernin concentration was not determined directly but was in the 1 mg/ml range. From this kinetic data pernin inhibition was estimated to be about 50% of the level of human plasma (approximately 1 mg/ml pernin diluted 9/10 compared with the 1/41 plasma dilution in the standard ATIII assay system). Heparin, a co-factor required for ATIII inhibition of thrombin, was not required for inhibitory action by pernin.

Metal Binding Activity

Hi Trap® Chelating affinity columns (Amersham Pharmacia Biotech, 1ml size) were prepared according to the manufacturer's instructions. The columns were then charged with either 0.1M cupric chloride or zinc chloride before equilibrating in a buffer (0.050M sodium phosphate and 0.5M sodium chloride containing 0.5mM imidazole, pH 7.0). Protein samples purified using CsCl centrifugation were suspended in this buffer and applied to the column using a chromatographic system (Econo System, Bio-Rad Laboratories, USA). Following washing of the column for 5 mins with buffer during which no protein appeared in the eluate, a linear gradient over 20 min at 1 ml/min was used to develop the column using buffer with the imidazole concentration at 100mM from 0-100%. The protein eluted into the gradient being retained longer on the copper chelation column than the zinc. The absorption of the eluate was monitored at 254nM.

Divalent metal ion content of the CsCl purified protein was determined by dissolving the protein in water at 10 mg/ml and analysing metal content by both atomic absorption and plasma emission spectrometry by comparison with a water blank. There was no significant divalent cation content in the protein purified by this method. However, purification by other methods not employing chaotropic agents like CsCl, the high content of histidine coupled with acidic amino acid residues and the likely origin of this protein from a SOD precursor, points to pernin having endogenous metal ions as part of its native structure.

Gene Sequencing Method

A suite of non-specific primers called pUZ5 was synthesised by Gibco-BRL for the initial sequencing based on the N-terminal sequence of pernin. The general formula was:

GAY GGN GAR CAR TGY AAY GAY GGN CAR AA

Where Y represents a pyrimidine base, R represents a purine base and N represents any one of the four nucleotide bases. Sequencing was done, initially using pUZ5 and an oligo-dT based "bottom stand" primer from PCR amplified cDNA. Sequencing was done by dye-termination cycle sequencing using "BigDye" prism technology (Applied Biosystems Incorporated, USA) according to their instructions.

Products were resolved on an ABI 377 automated sequencer. Following the initial sequencing of approximately 500 base pairs pernin-specific primers were constructed and used to complete the sequencing of the pernin gene.

This provided the following:

START
(AUG)

GATGGGGAGCAGTGTAACGATGGGCAGAACAAAGATGACCACCATGACGACCACCACGATGATCA
CCATGACGACCATGATGATGATGATGAAACAATGCACTATGCCCAGTGTGAAATGGAACCAAACC
CTCATATGGCTAGCAGCCTTCACCACCATGTCCATGGCAGCATAGAGTTGTCACAGAAGGGTCAT
GGAGCTGTTTATCTAGAACTTCATCTTGTCGGATTCAACACAAGTGAAGACCATGACGACCACCA
TCATGGACTTCATCTGCACATGCTTGGTGACATGTCAGCAGGTTGTGATTCTATTGGCGAACTGT
ACAATGCTCACCCAGAAAAACATGCTGACCCTGGTGACCTCGGTGACCTGGTTGACGATGATAGG
GGCGTGGTTAATGAAGTTCATCATTATGCTTGGTTGGACATTGATGGTACAGCACCAAACACCGA
AGCTCTCATTGGACACTCAATGACTATTTTACAAGGGAGTCACACCGATGCTGATACCCCAGCCA
GTAGAATCGCCTGTTGTGTTATTGGTCATGGAAAAGCTCGCCCAGAAACAGCAGCTGCTCTACAT
CACGAGCTAGAGGAAGATAAAACTGAGCATTATGCCCATTGTGACGTAAGATCTAATACACACCA
ACCAAAGGCTCTTCATCATCATGTCCACGGAACCATCGATTTCAAACAAGTTGGTTATGGTGACC
TTGAAGTGTCCTACCATTTAGAGGGATTTAATGTAAGTGATGACCACAAAGATCATCTCCATGAC
GTACAGATCTACGCCAACGGTGACCTGACCAGTGGATGTGATAACCTCGGTGCTAAATATGATCC
TCATGAAGATTACCACAGTGAGTTGGGTGATCTAGGAGATATTCACGATGATGACCATGGCGTTG
TCAATGAAAGCCACAGATATTCCTGGATCAATATCTTCGGTGATGACAGTGTCCTGGGACGTTCT
ATTGCCATTCACCAAAGAGACCATCTTCATAAAAGTGCCAAAATTGCCTGTTGTGTCATAGGACG
TGGACAGAGCCATCCAGAAATTGTTCACAGAGCTAAATGTGTTGTCAGACCTAATACAGAATCTA
CTGGTTTACATCACCATGTCTCTGGTTCTATAACATTCGAACAGACCCCTGGAGGATCAACACAT
ATGACGGCTGATCTCAAAGGATTTAACGTTAGTGAGGACTTGTCACATCATCGTCATGGTGTGCA
GCTCCATGAATGGGGAGATATGTCCCATGGCTGTCACTCCTTAGGCAGAATGTACCATGGTCATG
ATGATGCTCATGACCCCAAAAGACCTGGTGACCTTGGTGATGTTATAGATGATTCCCATGGCATC
GTTCATTCAACTAGAACCTTTGATCATCTTAATGTTGAAGATCTTAACGCACGTTCCCTTGTGAT
TATGCAGGGCGGACATGAGGTCGAGAGTGAGAGGGTTGCTTGCTGTGTTATAGGACGGGCA**TGA**A
TAACCTCACTAGAGTGACTTTGTCTAACATGACAATTAACAATTGTATAACTTCGCTAAAA**AATA**
AAACAATGACACAATGNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

C. Discussion

The present invention is a novel protein obtainable from *Perna canaliculus*, the New Zealand green-lipped (Greenshell™) mussel. The protein appears to be able to self-aggregate in structures resembling small virus like particles (VLPs) approximately 25 nm in diameter but lacking any nucleic acid. The protein was found in extracts of whole mussels and appears to be the predominant protein in haemolymph. The molecular weight of the protein was estimated to be 75 kDa by PAGE and inferred to be 55 kDa from its polynucleotide encoding sequence but, because of its ability to aggregate, the protein can be sedimented by ultracentrifugation in a short time (e.g. 40 minutes at 250,000 *g*) whereas the monomeric protein would not. Each ml of haemolymph yields, on the average, about 2 mg of pernin. Haemolymph is easily obtained by withdrawing fluid from the posterior adductor muscle of the shellfish which can yield up to 5 ml without obvious harm; it is not necessary to kill the mussel. The haemolymph obtained not only contains high levels of pernin but is quite free of contaminating materials, particularly compared with whole mussel extracts, so purification of pernin is simple. For highly pure preparations of pernin, ultracentrifugation is followed by isopycnic banding in a suitable density gradient medium such as CsCl.

The sequence of the N-terminus of pernin suggested that the protein might have anti-thrombin activity. This was demonstrated in kinetic assays on purified pernin. Since thrombin is a serine protease, pernin also acts as a serine protease inhibitor.

Comparison of the sequences obtained from several cleavage fragments against amino acid sequences in a computer database suggest that in addition to the anti-thrombin activity of pernin, the protein may also possess other activities. One of these is the ability to bind divalent cations such as Zn^{2+} and Cu^{2+}.

INDUSTRIAL APPLICATION

Because of its anti-thrombin activity pernin is potentially useful as an anti-coagulant agent. Thrombin normally acts as a protease which converts fibrinogen in the blood to fibrin. Blood coagulation is counteracted by inhibitors, normally anti-thrombin III (ATIII); pernin has also been shown to inhibit thrombin activity in an ATIII assay system. In contrast to ATIII, whose action is accelerated by the presence

336906

of heparin (a sulphated mucopolysaccharide) pernin does not require heparin as a co-factor.

The pernin protein from *P. canaliculus* may thus have value as a pharmaceutical. Since it is active as an anticoagulant in its native state it may also be useful as a natural therapeutic agent or health supplement. It is readily obtained as a natural product in high concentrations from mussel haemolymph. To obtain a highly pure preparation it is necessary only to remove haemocytes by centrifugation (or any other suitable method) followed by either ultracentrifugation (since pernin forms aggregates which readily sediment) and resuspension, isopycnic banding in a suitable medium such as CsCl, exclusion filtration through a suitable membrane which retains pernin, or chromatography through a medium such as controlled pore glass of suitable porosity. The result is a highly pure preparation of pernin.

The mussel *P. canaliculus* produces large amounts of the protein naturally, with little cost or effort involved in production, processing or purification.

Those persons skilled in the art will understand that the above description is provided by way of illustration only and that it is not to be regarded as limiting the scope of the invention.

REFERENCES

Layne, E. (1957). Spectrophotometric and turbidometric methods for measuring proteins, *Methods in Enzymology* **III**, 447.

Scotti, P.D. (1985). The estimation of virus density in isopycnic cesium chloride gradients. *Journal of Virological Methods* **12**, 149.

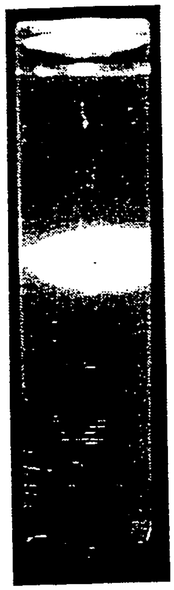
RUSSELL McVEAGH WEST WALKER

per [signature]

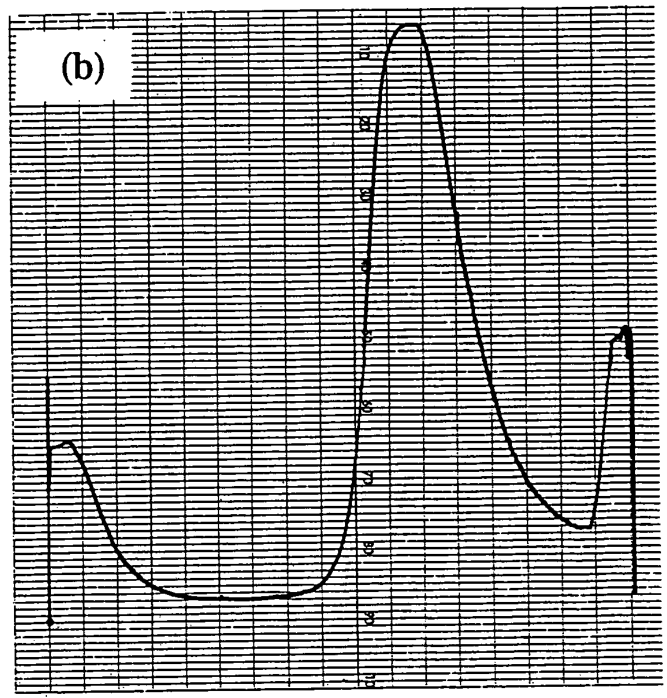
ATTORNEYS FOR THE APPLICANT

Figure 1

(a)



(b)



(c)

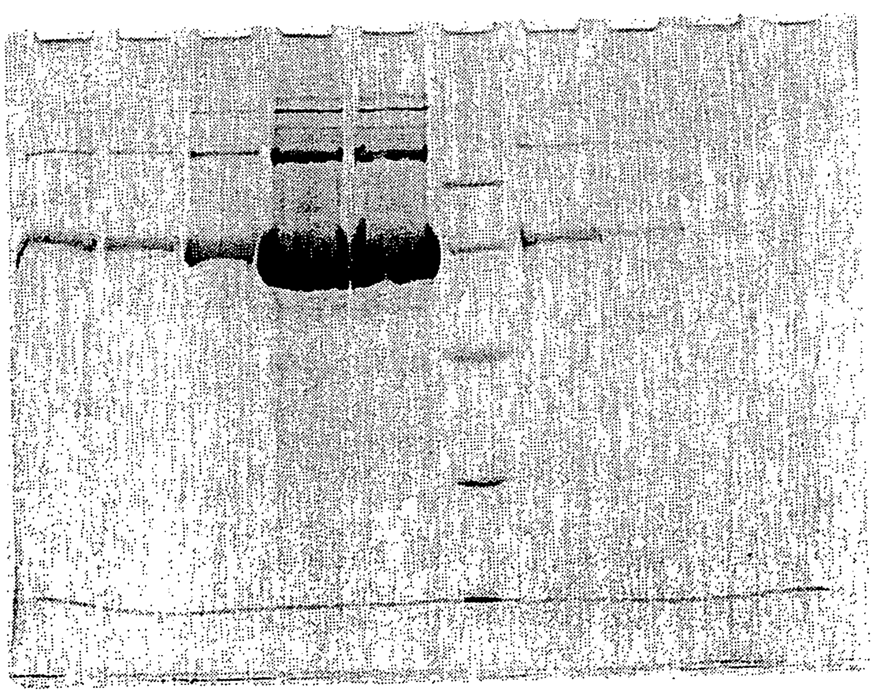


Figure 2

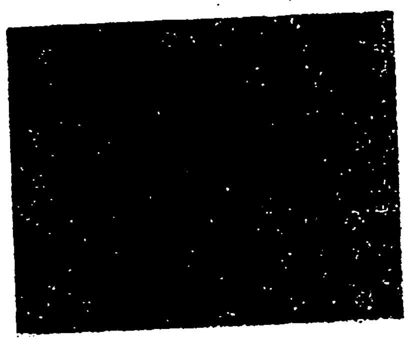


Figure 3

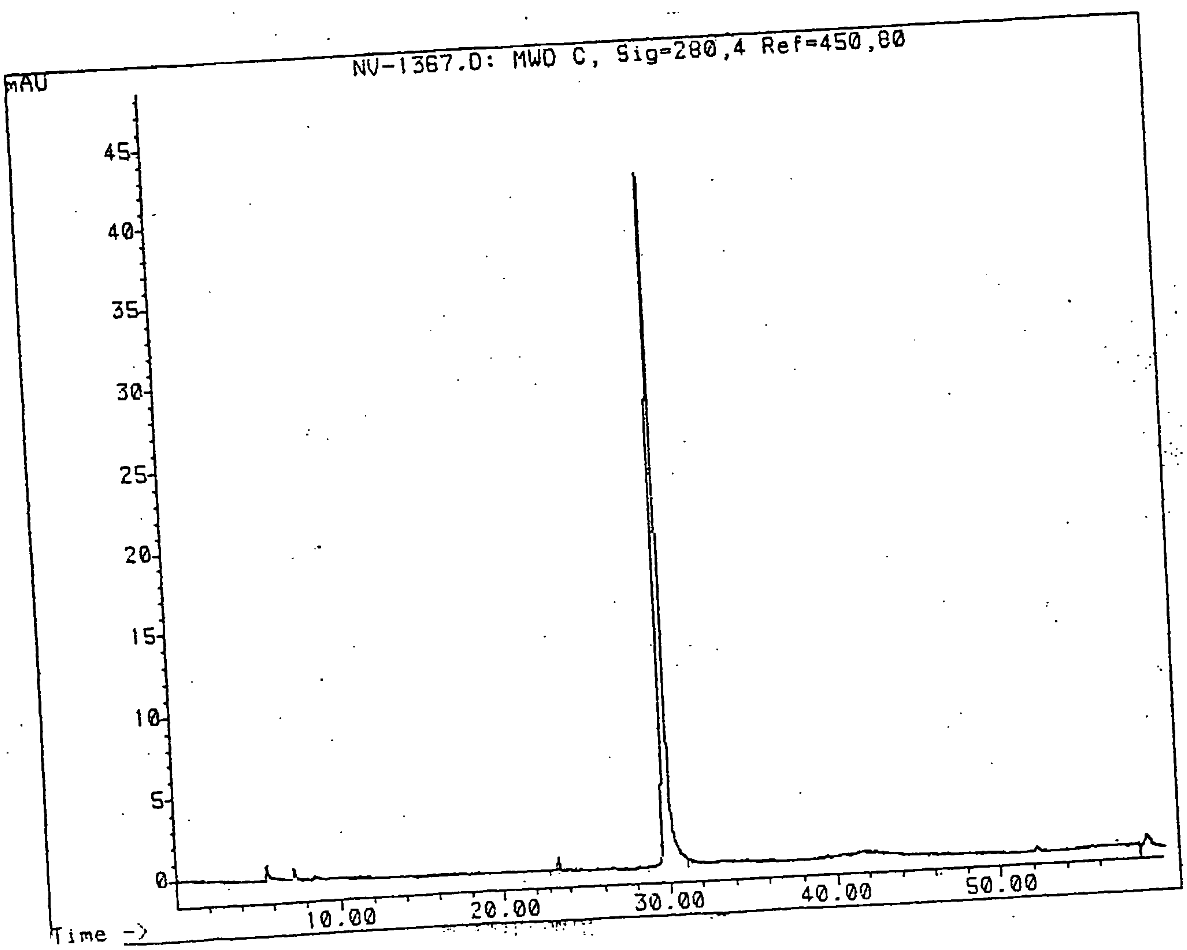


Figure 4a

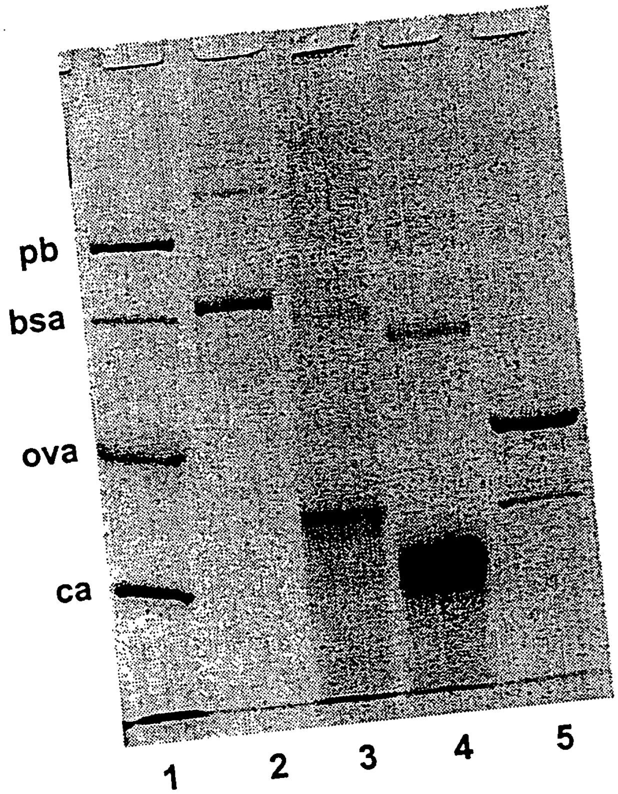


Figure 4b

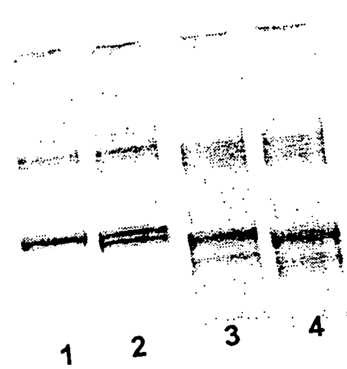
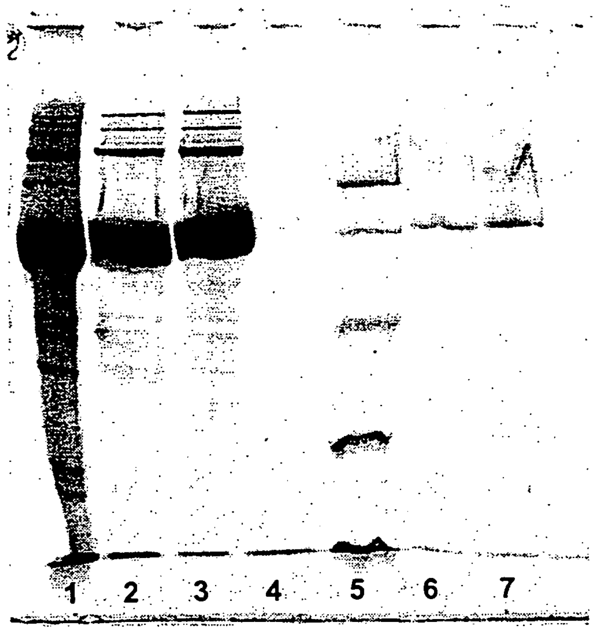
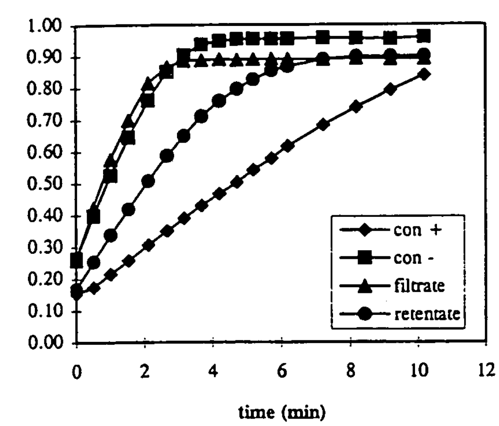


Figure 5

(a)


1
2
3
4
5
6
7

(b)


1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
od405
0
2
4
6
8
10
12
time (min)
con +
con -
filtrate
retentate